Deeper insight into the structure of the anaerobic digestion microbial community; the biogas microbiome database is expanded with 157 new genomes

Treu, Laura; Kougias, Panagiotis; Campanaro, Stefano; Bassani, Ilaria; Angelidaki, Irini

Published in:
Bioresource Technology

Link to article, DOI:
10.1016/j.biortech.2016.05.081

Publication date:
2016

Document Version
Peer reviewed version

Deeper insight into the structure of the anaerobic digestion microbial community; the biogas microbiome database is expanded with 157 new genomes

Laura Treu[1], Panagiotis G. Kougias[1*], Stefano Campanaro[2], Ilaria Bassani[1], Irini Angelidaki[1]

[1]Department of Environmental Engineering, Technical University of Denmark, Kgs. Lyngby, DK-2800, Denmark

[2]Department of Biology, University of Padua, Via U. Bassi 58/b, 35121 Padova, Italy

[*]Corresponding author: Panagiotis G. Kougias, Department of Environmental Engineering, Technical University of Denmark, Bld 113, 2800 Lyngby, Denmark. E-mail address: panak@env.dtu.dk, Tel.: +45 45251454

Abstract

This research aimed to better characterize the biogas microbiome by means of high throughput metagenomic sequencing and to elucidate the core microbial consortium existing in biogas reactors independently from the operational conditions. Assembly of shotgun reads followed by an established binning strategy resulted in the highest, up to now, extraction of microbial genomes involved in biogas producing systems. From the 236 extracted genome bins, it was remarkably found that the vast majority of them could only be characterized at high taxonomic levels. This result confirms that the biogas microbiome is comprised by a consortium of unknown species. A comparative analysis between the genome bins of the current study and those extracted from a previous metagenomic assembly demonstrated a similar phylogenetic distribution of the main taxa. Finally, this analysis led to the identification of a subset of common microbes that could be considered as the core essential group in biogas production.

Keywords

Biogas; microbiome; metagenome; assembly; binning

1. Introduction

Biogas production is a striking technology for sustainable generation of renewable energy. The produced biogas is derived as a result of the anaerobic decomposition of organic matter via a biological process mediated mainly by a complex consortium of bacteria and archaea (Luo et al., 2015). Despite the fact that this technology is well established, considering the proliferation of the biogas plants worldwide, fundamental aspects related to the microbiology of the process is still unclear.

In the cited literature the composition of biogas-producing microbial communities has been generally determined via construction of 16S-rRNA clone libraries and subsequent analysis of 16S-rRNA amplicons (De Francisci et al., 2015; Kröber et al., 2009; Luo et al., 2015). The taxonomic assignment of the microbial species was commonly based on sequence similarity search against reference 16S rRNA sequences deposited in public databases. Even in the shotgun sequencing studies, most of the reference genomes used for profiling the composition of the microbial communities are isolates from various environments different from the anaerobic digestion system, while it is known that only a small fraction of microorganisms have been cultivated (Albertsen et al., 2013; Hugenholtz, 2002). Therefore, even if the phylogeny of these genomes is related to the ones found in biogas communities, it is uncertain whether they serve the same function during the anaerobic digestion (AD) process. Moreover, by profiling only phylogenetic marker genes, such as the 16S rRNA gene, it is impossible to acquire insights on the community's functional capabilities (Langille et al., 2013), and thus fundamental information regarding essential roles of predominantly uncultivated microbes (e.g. symbiotic or competitive behavior) in the formation of a collective network are limited (Tyson et al., 2004). Another aspect of particular attention is the definition of a core microbiome in biogas production. Riviere et al., (2009) demonstrated that in sludge digesters there is a fraction of phylotypes that are always present constituting the common prokaryotic community, while another fraction of phylotypes are site specific. Nevertheless, such information are lacking in cases of biogas reactors treating agricultural and industrial residues. It is imperative to extend the analysis of the core microbiome at genomic level in AD systems elucidating the genome structure of the stable taxa and of those specific of different operational conditions.

The advancement of sequencing technologies and bioinformatic tools allow nowadays a deep characterization of complex communities, such as the one of the biogas microbiome. Therefore, in the cited literature the number of metagenomic analyses, even without performing assembly or binning processes, is increasing. Currently, most of the metagenomic studies on anaerobic digesters determined the functional properties of the microorganisms using non-assembled short reads (Eikmeyer et al., 2013), or in others works the gene finding was achieved using few number of short scaffolds (Schlüter et al., 2008; Stolze et al., 2015; Wirth et al., 2012). Bremges et al., (2015) assembled the metagenome of a single agricultural production-scale biogas facility and managed to reconstruct most of the genes involved in methane metabolism.

In our previous work, it was demonstrated that by assembling the shotgun metagenome sequences and following a binning strategy, it was possible to dissect the bioma of multiple thermophilic biogas reactors treating manure-based substrates (Campanaro et al., 2016). In this approach, de novo assembly procedure can be applied to analyze complex microbial communities generating a large set of scaffolds, which can be subsequently classified in single biological entities with a procedure named binning. This classification can be performed with different strategies, but the most innovative is based on the rationale that in different environmental conditions one bacterial species can be present at different relative abundances, consequently scaffolds belonging to the same genome change their coverage concertedly and they can be attributed to the same microbe (Albertsen et al., 2013; Nielsen et al., 2014). The results led to the identification of 106 microbial genomes (Genome Bins, GBs), and a conservative estimation indicated the presence of more than 450 microorganisms in the biogas microbial community. This estimate was derived considering that approximately

70% of the assembly could not be assigned to a specific GB. Moreover, this argument was further reinforced as during the assembly process, the reads belonging to the least abundant microorganisms were discarded.

This study is a continuation of the previous work aiming to further elucidate the biogas microbial community by enriching the biogas microbiome database with reference genomes present in anaerobic digesters. The samples were obtained from mesophilic and thermophilic continuous reactors used to upgrade and enhance biogas production via hydrogen assisted methanogenesis. The microbial community found in the current study was compared with the corresponding one of our previous assembly. This allowed the determination of similarities and differences among the microbiota and the identification of a potential existence of common microbes that can serve as the core essential group for biogas production.

2. Materials and methods

2.1 Reactor configuration and sample collection

Samples were obtained from the secondary reactor of a serial configuration operating either in mesophilic (35±1 °C) or thermophilic (55±1 °C) conditions. The collection of the samples was performed once the reactors were operating under steady state conditions (i.e. after a period of 3 Hydraulic Retention Times) before and after H_2 addition to ensure representative process conditions and microbial community stability. As the upgrading process occurred in the secondary reactor of the serial configuration, only samples from the secondary stage were analyzed. Each configuration was comprised by two Continuously Stirred Tank Reactors (CSTR) connected in series with volume ratio between the primary/secondary reactor equal to 0.75. For the mesophilic

conditions, the Hydraulic Retention Time (HRT) of the primary and secondary reactors were 25 and 33 days, respectively, while the corresponding HRT for the thermophilic setup were 15 and 20 days, respectively. The primary reactor of each set was serving as conventional biogas producing digester fed with cattle manure. The characteristics of the manure used as substrate are given in Table 1. The digestate of the primary reactor along with external H_2 gas were introduced to the secondary reactor in order to upgrade the biogas quality by coupling the CO_2 contained in the biogas with the injected H_2. The H_2 flow rate and a detailed description of the reactor operation are described by Bassani et al., (2015).

2.2 DNA extraction and high throughput sequencing

Initially, each sample was filtered using a 100 μm nylon cell strainer filter in order to remove all the fibrous residues of animal nutrition present in the digested manure. Subsequently, the samples were centrifuges at 2500 g for 10 min in order to recover ~2 g of pellet. Genomic DNA was extracted using RNA PowerSoil® DNA Elution Accessory Kit (MO BIO Laboratories, Carlsbad, CA). NanoDrop (ThermoFisher Scientific, Waltham, MA) and Qbit fluorimeter (Life Technologies, Carlsbad, CA) were used to evaluate the quality and quantity of the extracted DNA. Metagenome sequencing was performed using Illumina NextSeq 500 desktop system and Nextera XT kit (Illumina, San Diego, CA) for library preparation (150+150 bp).

2.3 Metagenomic assembly and binning process

Trimmomatic software was used to filter the raw reads in FASTQ format and to remove the adaptors (Bolger et al., 2014). Overlapped paired-ends were merged using Flash (Magoč and Salzberg, 2011) using standard parameters, except from the maximum

overlap parameter, which was set to 150 bases. Assembly and binning strategy was performed using a previously established method (Campanaro et al., 2016); all the perl scripts used for binning were obtained from "http://www.biogasmicrobiome.com/" (binning process v1). For the metagenome assembly both paired-end reads and single-end reads (both those merged using Flash and those which only one end passed the filtering step) were used. Reads were assembled with CLC Genomics workbench v. 5.1 (CLC Bio, Aarhus, DK) using CLC's *de novo* assembly algorithm, using a kmer of 63 and a bubble size of 60. Scaffolds shorter than 1000 bp were discarded. Sequence data reported in this study have been submitted to the National Center for Biotechnology Information as part of the BioProject PRJNA283612. The raw sequence data were deposited at the Sequence Read Archive under the accession number SRP058235 (Table S1). This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession LSQX00000000. The version described in this paper is version LSQX01000000. Moreover, further information about the extracted Genome Bins is available at the genome database www.biogasmicrobiome.com. The applied binning strategy required the identification of the 107 essential genes previously proposed (Dupont et al., 2012) from the entire gene set predicted using Prodigal (run in metagenomic mode) (Hyatt et al., 2012). Identification of the essential genes was performed using HMMER3 (http://hmmer.janelia.org/) (Finn et al., 2011) with a set of Hidden Markov Models obtained from Albertsen et al., (2013). Scaffold coverage required in the binning strategy was determined by aligning the reads on scaffolds using Bowtie2 software (Langmead and Salzberg, 2012) and analyzing the "bam output file" with the genomecov software (Bedtools package, (Quinlan and Hall, 2010)) followed by "average_coverage_bedtools.pl" perl script. The number of essential genes identified on

each GB was used to predict the completeness and duplication level of the GBs by executing the script "determine_bins_completenes.pl". The phylogenetic assignment of each GB was performed with two independent methods and the higher common taxonomic assignment was used to ascribe a provisional name to the GBs. The first taxonomic assignment was performed with Phylophlan (Segata et al., 2013) and using as input all the protein sequences of each GB previously determined. Phylophlan taxonomic assignment (Dataset S1) was chosen to discuss the GBs classification. The second method was applied on all the scaffolds assigned to each GB by the Phylopythia software (McHardy et al., 2007). Genome contamination, which can inflate genome completeness estimates, was determined both by checking the number of essential genes present in more than one copy on a single GB. The GBs having completeness lower than 50% and/or contamination higher than 20% were used only for taxonomic discussion. The extracted GBs can be downloaded from http://www.biogasmicrobiome.com.

2.4 Comparison of Genome Bins between the different assemblies

Identification of the common GBs between the two metagenomic assemblies was performed by determining the Average Nucleotide Identity (ANI) on the protein-encoding nucleotide sequences. Initially, makeblastdb software was used to generate a database with the nucleotide sequences of the 106 GBs obtained from a previous binning process (Campanaro et al., 2016). Similarity search was performed using BLASTN with all the nucleotide sequences of the genes identified for each GB, by using $1E^{-5}$ as minimum threshold. By using the script "ANI_calculator_CS.pl" the BLASTN output was analyzed and the number of sequences having a match in the

database was calculated. The same script provided as output the ANI value for each GB. Two GBs were considered as belonging to the same species if at least 50% of the genes found a match and if the average nucleotide-level similarity was higher than 95% (Nielsen et al., 2014; Rodriguez-R and Konstantinidis, 2010). Krona tool (Ondov et al., 2011) was used to visualize the number of GBs belonging to the different taxonomic groups obtained from the two assemblies.

3. Results and Discussion

3.1 Assembly and taxonomic analysis of the metagenome

Illumina NextSeq 500 sequencing of bulked metagenomic DNA from biogas upgrading reactors samples generated 44.39 billion bp of sequencing data (shotgun reads). After removing contigs shorter than 1 kbp, primary assembly resulted in 409,516 contigs with a total assembly size of 1.37 billion bp (N50 5,274). Approximately 32.2% of the assembly (442,244,761 bp) was assigned to the extracted 236 GBs (Table S2).

Strict requirements were applied to the extracted GBs (i.e. completeness higher than 20% and contamination lower than 50%) in order to select the high quality level GBs. Taxonomic classification demonstrated that, by using Phylophlan software, only 13% of GBs were identified at genus level (30 GBs) and 3% at species level (8 GBs) (Dataset S1). For the remaining 84%, the taxonomic assignment was possible only at higher level, confirming that the majority of microorganisms involved in the AD process for biogas production were not extensively characterized. In particular it was possible to assign 7% of GBs only at the highest level (phylum) as best classification (Dataset S1). A more detailed description of the taxonomic classification is given in Supplementary Information.

Phylogenetic assignment results indicated *Firmicutes* as the most dominant phylum accounting for the 60% of the total community (141 GBs) (Fig. 1 and Supplementary Dataset S1). The dominance of *Firmicutes* in biogas reactors is in accordance with previous studies (Kröber et al., 2009; Schlüter et al., 2008) and can be attributed to their capability in polysaccharide and oligosaccharide degradation (Krause et al., 2008). Hydrolysis is the first step for efficient conversion of plant biomass to intermediate compounds which are further degraded into methane. It is well known that in diverse ecosystems degradation of cellulose (the most abundant plant polysaccharide) is performed by species affiliated to the phyla *Firmicutes* and *Fibrobacteres*. These species encode the cellulosome, one of the most known cellulases which are used both for adhesion and enzymatic purposes (Flint et al., 2008). A comparison with the analysis conducted on the full-length 16S rRNA genes (16S AFL) extracted from the same samples obtained in this study revealed a predominant role of this phylum, accounting for the 50% (mesophilic samples) and the 70% (thermophilic samples) of the sequences (Bassani et al., 2015). Among *Firmicutes*, *Clostridia* was the most prevalent phylogenetic class (131 GBs), followed by *Erysipelotrichi* (8 GBs) and *Negativicutes* (1 GB).

The second most abundant phylum was *Bacteroidetes* including 29 GBs (12% of the total community). The central role of *Bacteroidetes* together with *Firmicutes* confirms the outcomes of previous studies, which indicated that microbes belonging to these phyla are contributing to the decomposition of cattle manure, mainly consisting of plant biomass residues, in biogas reactors (De Francisci et al., 2015; Luo et al., 2015). According to the relative abundance of 16S AFL, *Bacteroidetes* accounted for the 30% of the mesophilic community and the 8% of the thermophilic (Bassani et al., 2015). The

vast majority of *Bacteroidetes* was assigned to *Bacteroidales* (27 GBs) and the remaining to *Flavobacteriales* (2 GBs).

Proteobacteria was the third most abundant phylum including 23 GBs and accounting for 10% of the metagenome. The most dominant class within this phylum was *Gammaproteobacteria* (8 GBs) followed by *Betaproteobacteria* (5 GBs), *Deltaproteobacteria* (3 GBs) and *Epsilonproteobacteria* (2 GBs).

Both *Spirochaetes* and *Synergistetes* accounted for the 3% of the total community (7 GBs each). Most of the GBs assigned to these phyla were identified at least at family level. Finally, phyla *Chloroflexi* (6 GBs), *Actinobacteria* (5 GBs) and *Tenericutes* (5 GBs) had approximately 2% occurrence in the microbial community, while *Verrucomicrobia* (3 GBs) occurred in 1% of GBs. The low abundance of *Chloroflexi* and *Actinobacteria* reveals that the structure of the microbial community populating biogas reactors treating agricultural and industrial residues is highly different compared to AD systems processing sludge and wastewater. It is well reported that microbes affiliated to these phyla dominate the community in activated sludge systems (Albertsen et al., 2015; Wang et al., 2014), as they are aerobic or facultative anaerobic microorganisms, which grow in the influent feed. According to the MiDAS database (http://www.midasfieldguide.org/en/home.htm), 13 *Chloroflexi* (~9% of the total) and 29 *Actinobacteria* (~20%) were identified in the ecosystem of activated sludge and related wastewater treatment systems.Therefore, in strictly anaerobic environments, such as the biogas reactors, the proliferation of such microorganisms is not favoured and they remain in low relative abundance (3% *Chloroflexi* and 2% *Actinobacteria* in the present study). The least dominant phyla identified were *Acidobacteria*, *Fibrobacteres*, *Lentisphaerae*, *Planctomycetes* and *Thermotogae* which were represented by 1 GB each

(<1%).

The archaeal community presented a narrower phylogenetic diversity compared to the bacterial one. Within phylum *Euryarchaeota*, 3 GBs belonged to *Methanomicrobia* and 1 GB to *Methanobacteria*, assigned to *Methanothermobacter*. Members of class *Methanomicrobia* were identified as genera *Methanoculleus* (2 GBs), and *Methanosarcina* (1 GB). It is known that *Methanoculleus* genus dominates the AD community in reactors processing manure substrates (Campanaro et al., 2016; Kröber et al., 2009) and increases in abundance when grown on carbon dioxide (CO_2) and hydrogen (H_2) as substrates (Bassani et al., 2015; Maus et al., 2015). These genera were frequently reported as part of the biogas microbiome, but they do not represent the entire archaeal fraction. This is proven by the presence of an unknown Archaea distantly related to the seventh order of methanogens, the *Methanomassiliicoccales*, which was recently reported (Campanaro et al., 2016). In the cited study, it was unclear if the presence of this species was determined by the inocula or by the specific growth conditions examined or, at the contrary, if it is common in the biogas microbiome. In the present study *Euryarchaeota* sp. DTU008 was independently binned and quantified suggesting the presence in the AD microbiome of a niche occupied by methylotrophic methanogens (Supplementary dataset S1). Its relevance in methanogenesis was confirmed and it represents a possible candidate to the core biogas microbiome.

3.2 Comparing the extracted Genome Bins from different metagenomic assemblies

The presence of a "core microbiome" in the AD community implies that part of the GBs can be found in nearly every biogas reactor (Riviere et al., 2009). Therefore, one of

the crucial steps in the analysis is the identification of the common GBs in the two metagenomes. As previously reported, comparative analysis performed through similarity search allows a tentative assignment of the GBs at species level (Rodriguez-R and Konstantinidis, 2010) and obviously this procedure can be extended to the identification of the same GB in multiple samples (Nielsen et al., 2014). Both these studies suggested an average sequence similarity of 95% as a reference threshold for assignment to the species level. Similarly, we used the same requirements to assign the GBs derived from different assemblies to the same species. The number of orthologous genes identified between two strains of the same species is more variable, but most of the strains belonging to one species share at least 50% of their gene content (Rodriguez-R and Konstantinidis, 2010). The 165 GBs recovered in the present assembly having completeness >50% and contamination <20% were examined considering ANI and shared gene content. By analysing the results obtained from BLASTN similarity search on the protein encoding genes, we found that 75 GBs recovered from the two metagenomic assemblies (i.e. the one presented in the current paper and the previous one reported by Campanaro et al., (2016)) had an average nucleotide identity equal to or higher than 95% and at least 50% of the genes in common (Fig. 2a). Interestingly, 68 of these GBs had an average identity higher than 99% suggesting that they also belong to the same strain. The ordering of the ANI values in descending order evidenced that the vast majority of the GBs were in the range between 98.4% and 100% (Fig. 2b). After the threshold of 98.4%, ANI sharply decreased to 93.1% without any intermediate value confirming the correctness of the 95% as threshold for the identification of the same species in two different assemblies. Therefore, considering the results at species level, the current microbial analysis includes 71% of the 106 GBs previously identified

(Campanaro et al., 2016) and allowed the identification of 91 new GBs widening our perspective on the AD microbial community.

Moreover, in case the comparison is extended to include also the GBs with "20%< completeness <=50%" and "50%> contamination >=20%", the GBs in common with the previous study were 77 and the newly identified were 159. By extending the analysis to the total 265 unique GBs, the common GBs extracted in both metagenomic analyses are 29%, while 11% are specific of the first and 60% of the second analysis (Table 2 and Fig. 1). It is evident that most of the methanogenic archaeal species (67%) were identified in both assemblies. Specifically, the GBs assigned as *Methanothermobacter* (1 GB) and *Methanoculleus* (2 GBs) were common, while 1 GB assigned to *Methanosarcina* was classified as new as it was different from the other *Methanosarcina* previously identified. This resilience in the composition of the archaeal community is due to the fundamental role covered by the archaea in the biogas microbiome. Indeed, the methanogens are involved during the last step of organic matter degradation chain and they use a very strict range of substrates for methanogenesis (i.e. CO_2, H_2, acetate and methylamines). This explains why the species constituting the bacterial community are deeply influenced by the composition of the influent substrate, while archaea are not. Since the composition of the archaeal community is quite similar in different biogas reactors, it could be assumed that this constitutes a possible explanation for their accurate characterisation at species level. On contrary, a large fraction of the bacteria remain still unknown. Moreover, syntrophic bacteria (*Syntrophobacterales* and *Synergistia*) are also more stable and resilient due to their specialised functional role in the AD system (Werner et al., 2011). On contrary, the dynamics of fermenting populations, such as *Clostridia* and *Bacteroidetes*, were

markedly different from those of the syntrophs because they rely more on redundancy to maintain the overall function of the community. Concordantly, the functional properties of the *Bacteroidetes* (determined considering the gene content), positioned these species at the top level in the funnel concept of the biogas food chain previously proposed (Campanaro et al., 2016). Additionally, the modification of the operation temperature of the reactors (i.e. mesophilic or thermophilic conditions) affected deeply the population of the higher-level fermenting bacteria explaining why only 17% of the *Bacteroidetes* and 14% of the *Proteobacteria* were common in the two assemblies. *Firmicutes* group is more complex because it includes both syntrophic species and others involved in degradation of complex polymers. Regarding all the *Firmicutes* it results that the 36% of them belong to the core microbiome, but among this phylum the percentage raise up to the 50% by analysing separately the *Syntrophomonadaceae* family. This result can be explained considering their syntrophic functional specialization together with methanogens. To summarise, it can be concluded that independently from the operational conditions the biogas microbial community, as described till now, is composed by a number of recurrent phylotypes and by several other rare phylotypes.

3.3 Operational conditions influencing the identification of new GBs.

As previously described, this high number of newly extracted GBs is mainly attributed to the completely different conditions of the biogas upgrading reactors from which the samples were obtained compared with our previous metagenomic assembly. For example, the operating mesophilic conditions of the reactors used in the current study favoured the growth of specific microbes that were not identified in the previous assembly performed only in samples from conventional thermophilic biogas reactors.

The graphical representation of the extracted GBs from the two different assemblies is illustrated in Supplementary Information (Supplementary Fig. S1 and Fig. S2). The taxonomic analysis showed that these microorganisms were belonging to the phyla *Acidobacteria* (1 GB), *Chloroflexi* (6 GBs), *Fibrobacteres* (1 GB), *Lentisphaerae* (1 GB), *Planctomycetes* (1 GB) and *Verrucomicrobia* (3 GBs) (Fig. 1, Supplementary Fig. S1 and Fig. S2). Most of the above microorganisms are reported to be present in AD environments having functional roles related to conversion of organic matter, aromatic compound and acetate fermentation (Krakat et al., 2011). Moreover, previous studies using 16S rRNA gene amplicon sequencing have reported the presence of *Verrucomicrobia* sp. in biogas reactors (Han et al., 2016). Nevertheless, due to the absence of reference genome were not able to explain their functional properties. It should be highlighted that this is the first work which identifies two nearly complete and one partially complete genome sequence of *Verrucomicrobia* sp., which will further assist in unraveling their role in the biogas system. Moreover, the addition of H_2 was the main determinant of the proliferation of syntrophic bacteria belonging to *Syntrophomonadaceae* family. Additional variations in microbiome were identified as a consequence of the different operational conditions, i.e. temperature and H_2 addition. The different conditions resulted in increased occurrence of *Bacteroidetes*, whose occurrence in biogas upgrading reactors doubled compared to the corresponding one in conventional thermophilic biogas reactors. On the contrary, the operational conditions did not influence the share of *Firmicutes* in the microbiome but only increased the number of the extracted GBs. This evidences the consistent role of these microorganisms in the AD process. In fact, it is well known that members of these phyla pose diverse metabolic capabilities (e.g. production of volatile fatty acids) during

fermentation and thus enabling metabolic flexibility upon process disturbances (Krakat et al., 2011), such as the H_2 addition.

4. Conclusions

The current metagenomic analysis led to the most accurate, till now, description of the biogas microbiota identifying 236 genome bins. A comparative study demonstrated the existence of a potential core essential microbial group in biogas production. This microbial group is present independently from the operational conditions and is composed by several recurrent phylotypes including *Methanoculleus*, *Methanotermobacter*, *Synthrophomonas* and *Proteobacteria*. Composition of archaeal community was found to be resilient, while bacterial community was more diverse due to higher functional variability. These findings represent an important basis for future metatranscriptomic and metaproteomic studies and will assist to optimize the biogas production process.

Acknowledgement

This work was supported by the Danish Council for Strategic Research under the project "SYMBIO – Integration of biomass and wind power for biogas enhancement and upgrading via hydrogen assisted anaerobic digestion", contract 12-132654. Illumina sequencing was performed at the Ramaciotti Centre for Genomics (Sydney, Australia). The metagenomic assembly and the extracted genome bins can be downloaded at www.biogasmicrobiome.com.

References

1. Albertsen, M., Hugenholtz, P., Skarshewski, A., Nielsen, K.L., Tyson, G.W., Nielsen, P.H., 2013. Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. Nature biotechnology. 31, 533-538.
2. Albertsen, M., Karst, S.M., Ziegler, A.S., Kirkegaard, R.H., Nielsen, P.H., 2015. Back to Basics–The Influence of DNA Extraction and Primer Choice on Phylogenetic Analysis of Activated Sludge Communities. PloS one. 10, e0132783.
3. Bassani, I., Kougias, P.G., Treu, L., Angelidaki, I., 2015. Biogas upgrading via hydrogenotrophic methanogenesis in two-stage Continuous Stirred Tank Reactors at mesophilic and thermophilic conditions. Environmental science & technology. 49, 12585-12593.
4. Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics, 2114–2120.
5. Bremges, A., Maus, I., Belmann, P., Eikmeyer, F., Winkler, A., Albersmeier, A., Pühler, A., Schlüter, A., Sczyrba, A., 2015. Deeply sequenced metagenome and metatranscriptome of a biogas-producing microbial community from an agricultural production-scale biogas plant. Gigascience. 30.
6. Campanaro, S., Treu, L., Kougias, P.G., Francisci, D., Valle, G., Angelidaki, I., 2016. Metagenomic analysis and functional characterization of the biogas microbiome using high throughput shotgun sequencing and a novel binning strategy. Biotechnology for biofuels. 9, 1-17.
7. De Francisci, D., Kougias, P.G., Treu, L., Campanaro, S., Angelidaki, I., 2015. Microbial diversity and dynamicity of biogas reactors due to radical changes of feedstock composition. Bioresource technology. 176, 56-64.
8. Dupont, C.L., Rusch, D.B., Yooseph, S., Lombardo, M.-J., Richter, R.A., Valas,

R., Novotny, M., Yee-Greenbaum, J., Selengut, J.D., Haft, D.H., 2012. Genomic insights to SAR86, an abundant and uncultivated marine bacterial lineage. The ISME journal. 6, 1186-1199.

9. Eikmeyer, F.G., Rademacher, A., Hanreich, A., Hennig, M., Jaenicke, S., Maus, I., Wibberg, D., Zakrzewski, M., Pühler, A., Klocke, M., 2013. Detailed analysis of metagenome datasets obtained from biogas-producing microbial communities residing in biogas reactors does not indicate the presence of putative pathogenic microorganisms. Biotechnol Biofuels. 6, 49.
10. Finn, R.D., Clements, J., Eddy, S.R., 2011. HMMER web server: interactive sequence similarity searching. Nucleic acids research, gkr367.
11. Flint, H.J., Bayer, E.A., Rincon, M.T., Lamed, R., White, B.A., 2008. Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis. Nature Reviews Microbiology. 6, 121-131.
12. Han, S., Liu, Y., Zhang, S., Luo, G., 2016. Reactor performances and microbial communities of biogas reactors: effects of inoculum sources. Applied microbiology and biotechnology. 100, 987-995.
13. Hugenholtz, P., 2002. Exploring prokaryotic diversity in the genomic era. Genome Biol. 3, 1-0003.8.
14. Hyatt, D., LoCascio, P.F., Hauser, L.J., Uberbacher, E.C., 2012. Gene and translation initiation site prediction in metagenomic sequences. Bioinformatics. 28, 2223-2230.
15. Krakat, N., Schmidt, S., Scherer, P., 2011. Potential impact of process parameters upon the bacterial diversity in the mesophilic anaerobic digestion of beet silage. Bioresource technology. 102, 5692-5701.
16. Krause, L., Diaz, N.N., Edwards, R.A., Gartemann, K.-H., Krömeke, H., Neuweger, H., Pühler, A., Runte, K.J., Schlüter, A., Stoye, J., 2008. Taxonomic composition and gene content of a methane-producing microbial community

isolated from a biogas reactor. Journal of Biotechnology. 136, 91-101.

17. Kröber, M., Bekel, T., Diaz, N.N., Goesmann, A., Jaenicke, S., Krause, L., Miller, D., Runte, K.J., Viehöver, P., Pühler, A., 2009. Phylogenetic characterization of a biogas plant microbial community integrating clone library 16S-rDNA sequences and metagenome sequence data obtained by 454-pyrosequencing. Journal of Biotechnology. 142, 38-49.
18. Langille, M.G., Zaneveld, J., Caporaso, J.G., McDonald, D., Knights, D., Reyes, J.A., Clemente, J.C., Burkepile, D.E., Thurber, R.L.V., Knight, R., 2013. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nature biotechnology. 31, 814-821.
19. Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nature methods. 9, 357-359.
20. Luo, G., De Francisci, D., Kougias, P.G., Laura, T., Zhu, X., Angelidaki, I., 2015. New steady-state microbial community compositions and process performances in biogas reactors induced by temperature disturbances. Biotechnology for biofuels. 8, 3.
21. Magoč, T., Salzberg, S.L., 2011. FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics. 27, 2957-2963.
22. Maus, I., Wibberg, D., Stantscheff, R., Stolze, Y., Blom, J., Eikmeyer, F.-G., Fracowiak, J., König, H., Pühler, A., Schlüter, A., 2015. Insights into the annotated genome sequence of Methanoculleus bourgensis MS2 T, related to dominant methanogens in biogas-producing plants. Journal of Biotechnology. 201, 43-53.
23. McHardy, A.C., Martín, H.G., Tsirigos, A., Hugenholtz, P., Rigoutsos, I., 2007. Accurate phylogenetic classification of variable-length DNA fragments. Nature methods. 4, 63-72.
24. Nielsen, H.B., Almeida, M., Juncker, A.S., Rasmussen, S., Li, J., Sunagawa, S.,

Plichta, D.R., Gautier, L., Pedersen, A.G., Le Chatelier, E., 2014. Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. Nature biotechnology. 32, 822-828.

25. Ondov, B.D., Bergman, N.H., Phillippy, A.M., 2011. Interactive metagenomic visualization in a Web browser. BMC bioinformatics. 12, 1.
26. Quinlan, A.R., Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 26, 841-842.
27. Riviere, D., Desvignes, V., Pelletier, E., Chaussonnerie, S., Guermazi, S., Weissenbach, J., Li, T., Camacho, P., Sghir, A., 2009. Towards the definition of a core of microorganisms involved in anaerobic digestion of sludge. The ISME journal. 3, 700-714.
28. Rodriguez-R, L.M., Konstantinidis, K.T., 2010. Bypassing cultivation to identify bacterial species. Issues.
29. Schlüter, A., Bekel, T., Diaz, N.N., Dondrup, M., Eichenlaub, R., Gartemann, K.-H., Krahn, I., Krause, L., Krömeke, H., Kruse, O., 2008. The metagenome of a biogas-producing microbial community of a production-scale biogas plant fermenter analysed by the 454-pyrosequencing technology. Journal of Biotechnology. 136, 77-90.
30. Segata, N., Börnigen, D., Morgan, X.C., Huttenhower, C., 2013. PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. Nature communications. 4.
31. Stolze, Y., Zakrzewski, M., Maus, I., Eikmeyer, F., Jaenicke, S., Rottmann, N., Siebner, C., Pühler, A., Schlüter, A., 2015. Comparative metagenomics of biogas-producing microbial communities from production-scale biogas plants operating under wet or dry fermentation conditions. Biotechnology for biofuels. 8, 14.
32. Tyson, G.W., Chapman, J., Hugenholtz, P., Allen, E.E., Ram, R.J., Richardson,

P.M., Solovyev, V.V., Rubin, E.M., Rokhsar, D.S., Banfield, J.F., 2004. Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature. 428, 37-43.

33. Wang, C., Zuo, J., Chen, X., Xing, W., Xing, L., Li, P., Lu, X., Li, C., 2014. Microbial community structures in an integrated two-phase anaerobic bioreactor fed by fruit vegetable wastes and wheat straw. Journal of Environmental Sciences. 26, 2484-2492.
34. Werner, J.J., Knights, D., Garcia, M.L., Scalfone, N.B., Smith, S., Yarasheski, K., Cummings, T.A., Beers, A.R., Knight, R., Angenent, L.T., 2011. Bacterial community structures are unique and resilient in full-scale bioenergy systems. Proceedings of the National Academy of Sciences. 108, 4158-4163.
35. Wirth, R., Kovács, E., Maróti, G., Bagi, Z., Rákhely, G., Kovács, K.L., 2012. Characterization of a biogas-producing microbial community by short-read next generation DNA sequencing. Biotechnol Biofuels. 5, 41.

Figures

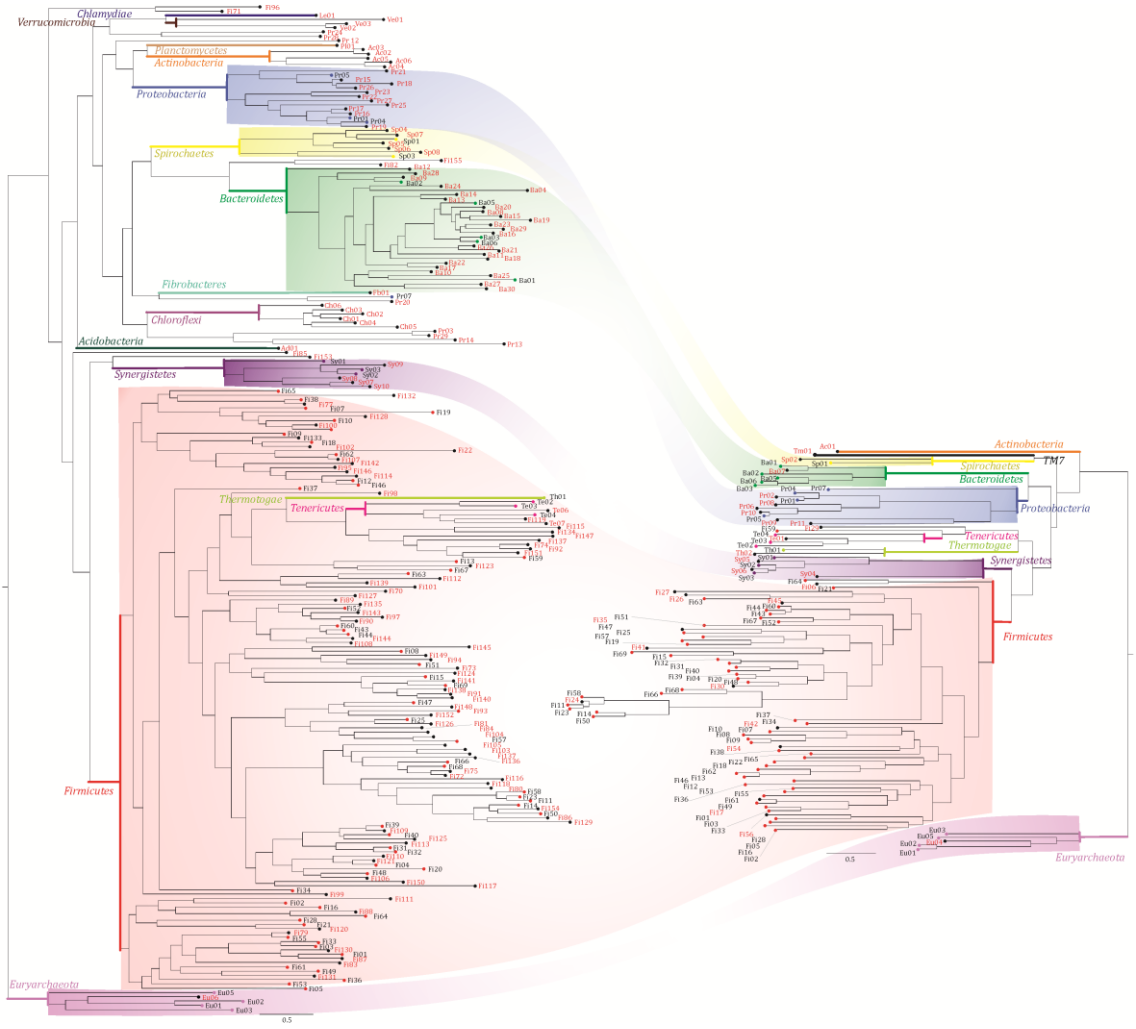

Fig. 1. Biogas phylogenetic trees. Comparison of high-resolution microbial trees with phylogeny and putative taxonomy, obtained using 400 broadly conserved proteins with PhyloPhlAn software. The two trees were built using FigTree and show the previously identified 106 GBs on the right and the newly binned 236 GBs on the left. Microbes are represented by small dots coloured according to the phylum when they were common between the two assemblies. Microbes represented by black dots and with ID in red colour text, were unique of one assembly.

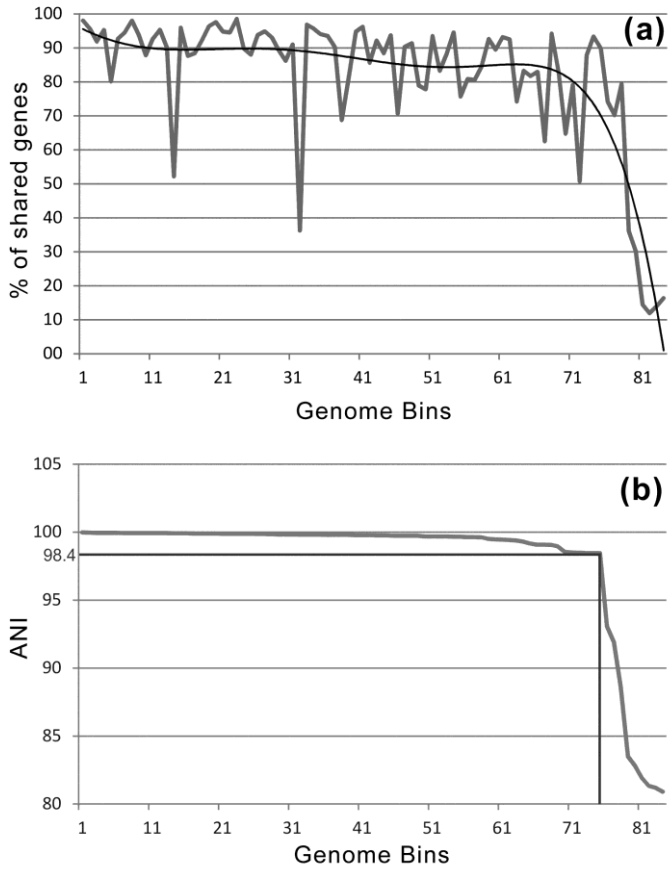

Fig. 2. Average nucleotide identity (ANI) and percentage of shared genes between the extracted Genome Bins from the two metagenomic assemblies. Only GBs with completeess >50% and contamination <20% were considered.

Table 1. Chemical composition of cattle manure used in the experiment

Parameter	Unit	Values
pH	-	7.44±0.01
Total solids (TS)	g/L	47.40±1.86
Volatile solids (VS)	g/L	34.56±1.40
Total Kjeldahl Nitrogen (TKN)	g-N/L	3.03±0.10
Ammonium Nitrogen (NH_4^+)	g-N/L	2.07±0.01
Total Volatile fatty acids (VFA)	mg/L	6831±477
Acetate	mg/L	4151±394
Propionate	mg/L	1421±67
iso-butyrate	mg/L	142±1
Butyrate	mg/L	793±16
iso-valerate	mg/L	224±1
Valerate	mg/L	88±1
n-hexanoate	mg/L	12±1

Table 2. Results comprising the biogas microbiome total 265 unique GBs (www.biogasmicrobiome.com), the common GBs extracted in both metagenomic analyses ("core microbiome") and GBs specific of the previous and of the current assembly.

	Biogas microbiome	**Core microbiome**		**Previous assembly**		**Present assembly**	
Taxon	Total GBs	*GBs*	*%*	*GBs*	*%*	*GBs*	*%*
Firmicutes	154	56	36%	13	8%	85	55%
Syntrophomonadaceae	48	24	50%	0	0%	24	50%
Bacteroidetes	30	5	17%	1	3%	24	80%
Proteobacteria	29	4	14%	6	21%	19	66%
Synergistetes	10	3	30%	3	30%	4	40%
Spirochaetes	8	1	13%	1	13%	6	75%
Actinobacteria	6	0	0%	1	17%	5	83%
Chloroflexi	6	0	0%	0	0%	6	100%
Euryarchaeota	6	4	67%	1	17%	1	17%
Tenericutes (Firmicutes)	6	3	50%	1	17%	2	33%
Verrucomicrobia	3	0	0%	0	0%	3	100%
Thermotogae	2	1	50%	1	50%	0	0%
Fibrobacteres	1	0	0%	0	0%	1	100%
Acidobacteria	1	0	0%	0	0%	1	100%
Chlamydiae	1	0	0%	0	0%	1	100%
Planctomycetes	1	0	0%	0	0%	1	100%
TM7	1	0	0%	1	100%	0	0%
Total	265	77	29%	29	11%	159	60%